

THE EFFECT OF FOREKNOWLEDGE OF DEMAND IN CASE OF A RESTRICTED CAPACITY: THE SINGLE-STAGE, SINGLE- PRODUCT CASE WITH LOST SALES

J. Wijngaard

**SOM-THEME A: PRIMARY PROCESSES WITHIN
FIRMS**

Abstract

Foreknowledge of demand is useful in the control of a production-inventory system. Knowing the customer orders in advance makes it possible to anticipate properly. It is an important condition to produce and deliver the right quantity of the right product “just-in-time”. It reduces the need of safety stock and spare capacity. But the question of the effectiveness of foreknowledge is not an easy one. Having foreknowledge of the customer orders does not remove the demand uncertainty completely. The effect of foreknowledge has to be considered in a stochastic dynamic setting. The subject of this paper is the effect of foreknowledge in combination with a restricted production capacity. The lost-sales case is considered. The main result is that for high utilization rates and small forecast horizon, the inventory reduction due to foreknowledge is equal to $(1-\rho)h$, with h the forecast horizon

1 Introduction

Foreknowledge of demand is useful in the control of a production-inventory system. Knowing the customer orders in advance makes it possible to anticipate properly. It is an important condition to produce and deliver the right quantity of the right product “just-in-time”. It reduces the need of safety stock and spare capacity. But the question of the effectiveness of foreknowledge is not an easy one. Having foreknowledge of the customer orders does not remove the demand uncertainty completely. The effect of foreknowledge has to be considered in a stochastic dynamic setting. The early papers of Baker [1],[2], Baker and Peterson[3] and Blackburn and Millen [4],[5] on rolling schedules have stressed this point. To estimate the performance of a system that is controlled by a rolling schedule is very complicated. If there is no foreknowledge of the customer orders, demand may be modeled in a pure stochastic way, inventory level and state of production give a sufficiently rich description of the state of the system. The transition mechanism is rather simple. If there is foreknowledge of the customer orders, the forecasts have to be included in the state description, making the transition mechanism much more complicated. This is probably the main reason why the performance of such rolling schedule controlled systems with partial foreknowledge have got so little attention in the operations management literature.

This combination of foreknowledge and uncertainty in the use of rolling plans, without having tools available to analyze such systems has confused the early discussions on the performance of MRP-controlled systems. In the analysis of a reorder point controlled system, the models include the stochastic nature of the environment, in particular the demand. This is possible because the demand is assumed to be purely stochastic. In the description of the advantages of MRP and the

effect of its use on the performance of the system, uncertainty has been neglected generally. The analysis of the behavior of the system has been reduced to static calculations. Nervousness, the most visible result of uncertainty, has been acknowledged thereafter as a problem, and has been given attention (see Blackburn/Kropp/Millen[6]). It has also stimulated some research on the matter of safety time versus safety stock (see Whybark/Williams[16]). See Vollmann/Berry/Whybark[15] and Graves[10] for more references. See also Wijngaard/Wortmann[17]. Thorough analyses of the performance of MRP-controlled systems are still rare, however. Extensive simulation studies are necessary to learn about the performance of an MRP-controlled system. See, for instance, Strickwold[12]. But it is not easy to derive general insights from the simulation of complete, complex, realistic systems.

To acquire insight into the effect of foreknowledge on performance, it is necessary to investigate characteristic, simple situations. The results of De Bodt/Van Wassenhove [9], for instance, on the effect of uncertainty on the performance of using the Wagner-Whitin algorithm or the Silver-Meal heuristic leads to insights that can also be used in more complex situations. Other interesting studies on elementary production-inventory systems, aiming at insight in the effect of uncertainty are Buzacott/Shantikumar[8], on safety time versus safety stock, and Hariharan/Zipkin[11](also referring to Van der Veen[13],[14]), on the equivalence of the reduction of demand uncertainty and leadtime reduction. The book of Buzacott/Shantikumar[7] gives also some results on the effect of foreknowledge. This paper is meant to contribute to production-inventory theory in the same spirit.

The subject of this paper is the influence of a restricted capacity on the effect of foreknowledge. This effect is studied by considering a model with one standard product that is made to stock. The most common assumption in such a production inventory model is the backlogging assumption. This paper deals with the lost-sales

case however. That is because the approach followed here is slightly clearer for the lost-sales case. To keep the analysis feasible, it is assumed that production is immediate. In a subsequent paper the backlogging case is going to be considered, in combination with the multi-product case. The effect of a positive throughput time will also get attention then. Foreknowledge is modeled as a positive customer order leadtime. The customer orders are known h periods in advance. The effect of changes in h are investigated. Knowing the customer orders in advance makes it possible to anticipate on capacity problems. If future demand is higher than the immediately available capacity, foreknowledge can be used to build sufficient inventory *just-in-time*. Without foreknowledge working with inventory means that there is always inventory. The inventory is only used incidentally. It is *just-in-case* inventory. Foreknowledge makes the inventory more effective. It makes it possible to accept more orders and deliver them in time with less inventory. The topic of this paper is the question of how much. The main messages are that the concept *virtual inventory* is useful to determine the effectiveness of foreknowledge and that for high utilization rates the rate at which the inventory (necessary for a certain utilization rate) decreases with increasing h , is equal to $1-\rho$.

The next section describes the model more precisely and introduces the concept virtual inventory. Section 3 proves that for small h and high utilization rates, optimal usage of foreknowledge leads to an inventory reduction of $(1 - \rho) \cdot h$. This theoretical result is illustrated with some simulation results. For larger h , the effect is smaller. Section 4 investigates this by analysis and simulation. Section 5 gives conclusions and suggestions for further research.

The paper is a combination of analysis, illustration of analytical results by simulation and more systematic simulations to extend the analytical results. The simulation programs are available for public use on www.wyngaard.nl.

2 Controlling the virtual inventory

There is one standard product. Customer orders for that product arrive according to a given stochastic process. The size of the orders follows some given distribution function. The orders are known h periods in advance (same for all orders). Upon arrival, it is checked whether it is possible to accept the order and deliver it in time. If not, the order is rejected. Orders are delivered from stock. That can be stock that is already available when the order arrives or stock that is produced between the moment of arrival and the moment of delivery. All accepted orders are delivered in time. The production speed is fixed. Production is in runs with a fixed length (Q). The production becomes immediately available, during the run. As soon as the run is over, it may be decided to start a new run. The system is controlled by these production decisions and by yes or no accepting customer orders. The aim is to realize a pre-specified utilization rate with as few inventory as possible (minimal average inventory).

Foreknowledge gives the possibility to produce in advance. So, foreknowledge can be used to decouple production from delivery. The result is reserved inventory. The foreknowledge horizon h forms a buffer between demand and capacity. Because there is only one product and all the customers demand the same product, it is also acceptable to produce free inventory. In case of no foreknowledge, orders have to be delivered immediately and are rejected therefore if no inventory is available. The maximum allowed (free) inventory forms a buffer between demand and capacity and can be determined such that the average idle time of the production facility is equal to some pre-specified value. Full utilization requires a high buffer. In case of foreknowledge, the customer order leadtime contributes also to this buffer. The important state variable is the free inventory:

$$J(t) := I(t) - C(t),$$

with $I(t)$ the inventory at time t and $C(t)$ the total customer order backlog. Upon arrival it has to be checked whether it is possible to deliver in time. Timely delivery is possible if:

$$J(t) \geq -R.h + s,$$

with R the production speed and s the size of the arriving order. So, $J(t) + R.h$ functions as a kind of buffer content. It is called the *virtual inventory*:

$$V(t) := I(t) - C(t) + R.h.$$

That is, $V(t)$ is the inventory at time $t + h$ if the production is *on* during the whole time interval $[t, t+h]$. The most straightforward control rule accepts as much as possible:

$$\text{accept if } V(t) \geq s,$$

and quits production if $V(t)$ comes above some critical level B :

$$\text{keep starting production runs as long as } V(t) < B.$$

Such control rules are called *simple* and are denoted by $(0, B)$. The behavior of $V(t)$ under control rule $(0, B)$ is completely identical to the behavior of the inventory itself under the control rule $(0, B)$ in case of no foreknowledge (and corresponding starting

state). If the arrival patterns are the same, the same orders will be accepted and the same production pattern will arise.

This can be used to determine a lower bound for the effect of foreknowledge. The inventory at time $t+h$ is equal to :

$$I(t+h) = V(t) - L(t,t+h).R,$$

with $L(t,t+h)$ the idle time during the interval $[t,t+h]$. Let ρ be the pre-specified utilization rate and suppose that B is such that ρ is realized in the long run. Then the long term average inventory is $(1 - \rho).R.h$ smaller than the long term average virtual inventory. But the long-term *average virtual inventory* in case of foreknowledge h (customer order leadtime is h) is equal to the long-term *average inventory* in case of no foreknowledge. So, by restricting the attention to simple control rules, foreknowledge h leads to a reduction of $(1 - \rho).R.h$ of the average inventory. It is possible that there are better rules of course. So, it is just a lower bound for the effect of foreknowledge.

There are two important questions with respect to these simple control rules. In the first place the question whether a simple control rule does not lead to too late deliveries. In the second place the question whether the optimal control rule is also of this simple type. The next section considers these issues in a more formal way and shows that, under certain conditions, both questions can be answered confirmatory.

3 The optimality of simple control rules for small h .

This section considers the behavior of the system under a simple control rule more extensively. First it is going to be proved that applying a simple control rule $(0,B)$ in case of foreknowledge h does not lead to too late deliveries if $h \leq B$ (lemma's 1 and 2). Thereafter the optimality of simple control rules is considered (lemma's 3 and 4). For sake of convenience, we assume $R = 1$. This does not lead to any loss of generality. It is just a matter of measuring the production speed in the right units.

Lemma 1

Compare the cases with foreknowledge h_1 and foreknowledge h_2 . Both are controlled with a $(0,B)$ -rule. Let I_1 and I_2 be the starting inventories, with I_1 and I_2 , such that $I_1 + h_1 = I_2 + h_2$ and suppose there are no customer orders available at time 0. Apply the same inter arrival pattern and order sizes to both systems. The resulting patterns of production and acceptance are completely identical (the same orders are accepted and rejected and the production is on during the same intervals).

Proof

The proof follows directly from the fact that a simple control rule works on the virtual inventory. If $I_1 + h_1 = I_2 + h_2$ and if there are no customer orders at time 0, the starting virtual inventories for the two cases are equal. Because the same $(0,B)$ rule is applied in both cases, the virtual inventories remain equal and the same decisions are taken with respect to acceptance and production.

Lemma 1 implies that the case without foreknowledge ($h = 0$) and starting inventory I_1 leads for each order pattern (inter arrival times and order sizes) to the same pattern of acceptance as the case with foreknowledge h (> 0) and starting inventory $I_1 - h$ if in both cases control rule $(0,B)$ is applied. The control rule $(0,B)$ keeps the virtual inventory at the start of a production run between 0 and B . This is used to prove that applying $(0,B)$ for a case with foreknowledge $h \leq B$, does not lead to too late deliveries:

Lemma 2

Suppose control rule $(0,B)$ is applied in case of foreknowledge h , with $B > h$. Let the starting inventory be equal to $B - h$ and let there be no customer orders at time 0. Then the inventory itself remains nonnegative and all accepted orders can be delivered in time.

Proof

The inventory itself and the virtual inventory are coupled in the following way:

$$I(t+h) = V(t) - L(t,t+h),$$

with $L(t,t+h)$ the idle time during $[t,t+h]$. This expression follows from the assumption that all accepted orders are delivered in time. It is sufficient to prove that $V(t) \geq 0$ implies $I(t+h) \geq 0$. Suppose $V(t) = x$, $0 \leq x \leq B$. It lasts at least $(B-x)$ time units before the production stops. That means that the idle time during the interval $[t,t+B]$ is at most equal to x and since $h \leq B$, x is also an upper bound for the idle time during the interval $[t,t+h]$. That implies $I(t+h) \geq 0$.

Lemma 2 shows that simple control rules with $B > h$ are allowed indeed. The matter of optimality has yet to be considered. But first some simulation results are given to illustrate the foregoing results. The simulation results are for the case of Poisson arrival (arrival rate λ) and order size equal to 1. Results in terms of accepted and not accepted orders and average inventory are shown for different values of B , h , Q , various random number generators and various simulation run lengths. The starting inventory is in all cases equal to $B - h$. In case of a simulation run of T time units the system simulates $T + h$ time units. Accepted and not accepted orders are counted over the first T time units. To let the starting inventory have not too much influence on the average inventory, the inventory and stock-out during the first h time units is not taken into account in determining the averages. Table 1 gives some results for the case without foreknowledge ($h = 0$). It illustrates the influence of B on the utilization rate.

Table 1: The case $h = 0$, $Q = 1$ and $\lambda = 1$ (simulation run length = 1,000 time units, three different random number generators)

$Q = 1$, $h = 0, \lambda = 1$				
B	#accepted	#notaccept	Inventory	Util. rate
5	835	76	3.450	0.835
5	911	74	3.110	0.907
5	909	89	3.377	0.907
6	851	60	4.210	0.852
6	933	52	3.862	0.928
6	920	78	3.673	0.917

For $B = 6$ the utilization rate is about equal to 0.9. The simulation results show that 1,000 time units is far too less to get a proper estimate of the expected utilization rate.

This is not a problem since the results on the effect of foreknowledge are sample path independent. But to illustrate how the transient state goes to the stationary state, table 2 gives the results for simulation runs of 10.000 time units (same pseudo random generators).

Table 2: The case $h = 0$, $Q = 1$ and $\lambda = 1$ (simulation run length = 10,000 time units, three random number generators)

Q = 1 , h = 0, $\lambda = 1$				
B	#accepted	#notaccept	Inventory	Util. rate
6	9097	806	3.819	0.909
6	9171	692	3.710	0.917
6	9154	788	3.547	0.915

The results show how slow the transient state approaches the stationary state. The influence of the sample path has to be taken into account in interpreting the results that are going to be derived here. The influence of Q is illustrated in the next table.

Table 3: The influence of Q (simulation runs of 10.000 time units, one random number generator).

Q = 10 , h = 0, $\lambda = 1$				
B	#accepted	#notaccept	Inventory	Util. rate
6	9174	729	4.302	0.917
5	9103	800	3.989	0.910

Because of the continuous character of the production, Q has not much influence and the case $Q > 1$ will not get any attention in the rest of the paper. The table shows how the critical level B has to be reduced somewhat to compensate for an increase in Q .

The influence of λ is essential, because it influences severely the inventory that is necessary to realize a certain pre-specified ρ . The next table illustrates this. To realize the same pre-specified ρ , the critical level B has to be increased significantly.

Table 4: The influence of λ (simulation runs of 10.000 time units, one random number generator).

Q = 1 , h = 0				
$\lambda = 0.95$				
B	#accepted	#notaccept	Inventory	Util. rate
6	8806	602	4.021	0.880
10	9095	313	6.597	0.909

The next table shows the effect of foreknowledge for the case of $Q = 1$, $\lambda = 1$, $B = 6$. Applying (0,B)-rules in case of foreknowledge with $B < h$, leads to stock-outs. A column with the average stock-out is added therefore. Different random number generators are used to show the sample path influence.

Table 5: The influence of foreknowledge (simulation runs of 1.000 time units, three random number generators for $h = 0$ and $h = 5$, one random number generator for $h = 10$ and $h = 15$).

Q = 1, B = 6 $\lambda = 1$						
h	#accepted	#notaccept	Inventory	Util.rate		Estimate
0	851	60	4.210	0.852	0	
0	933	52	3.862	0.928	0	
0	920	78	3.673	0.917	0	
5	851	60	3.477	0.852	0	3.470
5	933	52	3.512	0.928	0	3.502
5	920	78	3.258	0.917	0	3.258
10	851	60	2.777	0.852	0.039	2.730
15	851	60	2.206	0.852	0.212	1.990

The last column gives the estimate of the inventory based on the inventory for $h = 0$ and the sample path utilization rate. The results confirm that the inventory reduction because of foreknowledge, using (0,B)-rules, is equal to $(1 - ur) \cdot h$, with ur the sample path utilization rate. That holds also if $h > B$, but then the (0,B)-rule leads to stock-outs. The next section discusses a possibility to adapt this rule to prevent stock-outs.

Now the point of optimality of (0,B)-rules is going to be discussed. First it is proved that in case of Poisson arrival and order size 1, the optimal control rule is of the (0,B)-type if there is no foreknowledge ($h = 0$).

Lemma 3

Let the arrival process be Poisson and the customer order size be equal to 1. Suppose there is no foreknowledge ($h = 0$). Let ρ be the pre-specified utilization rate

that has to be realized. There is a simple control rule that realizes p with the lowest possible average inventory.

Proof

Replace the problem by a problem where each accepted order gives a reward r , instead of having a pre-specified utilization rate. This is a stationary Markov decision problem with the inventory as (only) state variable. The optimal policy is also stationary. That means that the optimal policy is of the form:

- Start a new production run if $I(t) \in P$
- Accept the order if $I(t) \in A$

, with P and A subsets of $(-\infty, +\infty)$.

Standard reasoning for stationary Markov decision problems leads to the conclusion that $A = [1, \infty)$ and P is of the form $(-\infty, B]$. So the optimal policy for this problem is of the $(0, B)$ -type.

Now, consider the original problem with a pre-specified p . There is an r such that applying the optimal policy for the just introduced Markov decision process leads to this pre-specified p . This policy is also optimal for the lost-sales problem with pre-specified p . This completes the proof.

Lemma 3 shows that the optimal policy for the case $h = 0$ is a *simple* control rule. Lemma 2 shows that this rule can also be applied for the case $0 < h \leq B$. The rule is also optimal for that case, because of the coupling of $I(\cdot)$ and $V(\cdot)$. This leads to lemma 4:

Lemma 4

Suppose $0 < h \leq B$. The $(0,B)$ -rule that is optimal (= minimal average inventory) for the case without foreknowledge, is also optimal for the case with foreknowledge h (customer order delivery time = h).

Proof

Lemma 2 proves that policy $(0,B)$ is possible for the case with foreknowledge. Lemma 3 proves that the average virtual inventory is minimal for $(0,B)$. The pre-specification of the utilization rate implies that $(0,B)$ minimizes also the average inventory.

The four lemma's lead straightforwardly to the result that the effect of foreknowledge is less than or equal to $(1 - \rho)h$. It is only equal to that if the simple control rule that is optimal for the case without foreknowledge can also be applied in the case with foreknowledge. If this rule leads to stock-outs, adjustments are required. Such adjustments lead to increases in the average virtual inventory and therefore also to increases in the average inventory itself (by the coupling $I(t+h) = V(t) - L(t,t+h)$).

The proofs of the lemma's show that the assumption of Poisson inter arrival times and the assumption of order sizes equal to 1 are not strictly necessary. The equivalence of the case without foreknowledge and the case with foreknowledge is more general. The optimal policy for the case without foreknowledge minimizes also the average inventory for the case with foreknowledge as long as this policy does not lead to stock-outs. In case of more general inter arrival times, the optimal policy may have to take into account the time since the last arrival. But also in such cases one may expect the existence of some $B > 0$ such that the production has to be put on as long as the inventory is smaller than B , independent of the time since the last arrival. The optimal policy for such a case without foreknowledge is also optimal for the

corresponding case with foreknowledge $h \leq B$. Relaxation of the assumption of order sizes equal to 1, may make it attractive to skip certain incoming customer orders because they do not fit nicely in the available inventory and capacity. But it may be conjectured that it is still optimal to control the production by the application of some critical level B (start a new run if and only if the inventory $< B$).

4 More foreknowledge ($h > B$)

In case of more foreknowledge ($h > B$) application of the policy that is optimal for the case without foreknowledge, may lead to stock-outs. That is because the idle time during the interval $[t, t+h]$ may exceed $V(t)$ (see the proof of lemma 2). This can be circumvented straightforwardly by adding a condition for starting a new production run. This condition checks on the customer order backlog whether starting a new run may be postponed. Define:

$$\text{slack}(t) = \min_j \{I(t) - j + dd_j\},$$

with dd_j the due date of order j . The start of a new production run may not be postponed beyond $\text{slack}(t)$. Figure 1 illustrates the condition.

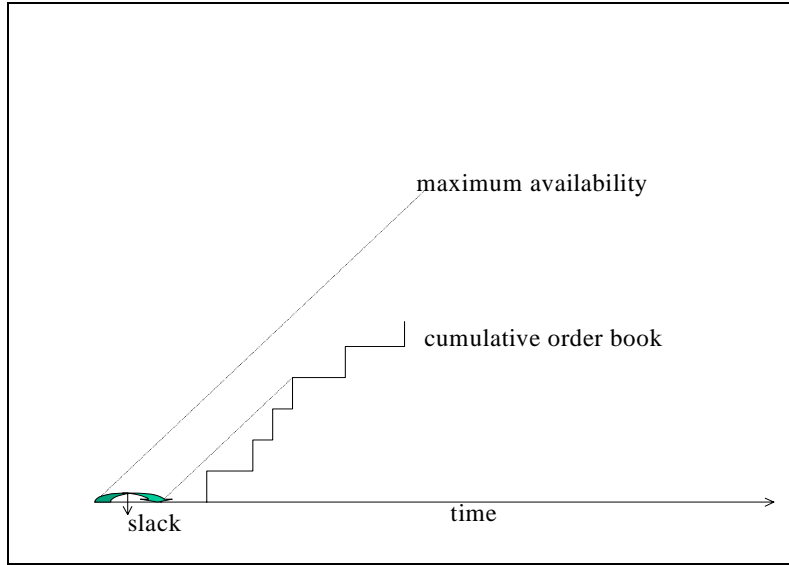


Figure 1: Checking the available slack

It may not be expected that the policy resulting from adding this production condition is optimal. But it is a reasonable policy and it is worth to investigate its performance. First some simulation results are given.

Table 6 The influence of checking the slack (simulation runs of 10,000 time units)

Q = 1, $\lambda = 1$						
h	B	#accepted	#notaccept	Inventory	Util.rate	Stockout
10	6	9097	806	2.931	0.909	0.016
10	6	9104	799	2.949	0.910	0
15	6	9097	806	2.571	0.909	0.108
15	6	9128	775	2.783	0.913	0
15	5.6	9097	806	2.566	0.909	0
20	6	9097	806	2.282	0.909	0.270
20	6	9170	733	2.636	0.917	0
20	5	9124	779	2.366	0.912	0
20	4.5	9109	794	2.326	0.911	0

As mentioned already, the policy without the check on available slack leads to stock-outs. Adding the extra condition removes the stock-outs again with as consequence a higher ρ and a higher average inventory. To compensate for the increase of ρ , the critical level B has to be reduced a little.

The policy resulting from checking the slack and adapting the production rule accordingly may not be expected to be optimal, although it is not easy to construct improvements. To acquire more insight in this problem it is useful to consider cases with more foreknowledge.

If there is sufficient foreknowledge it is possible to produce completely on order. That means that the critical level B can be put equal to 0. Without foreknowledge it is

necessary to use $B = 6$ to realize a utilization rate of (about) 0.910. In case of complete make-to-order, foreknowledge of (about) 33 time units is necessary to realize the same utilization rate. Table 7 illustrates this.

Table 7 The pure make-to-order case (simulation runs of 10,000 time units)

Q = 1, $\lambda = 1$				
B = 0				
h	#accepted	#notaccept	Inventory	Util.rate
33	9100	803	2.055	0.910
35	9128	775	2.122	0.913

If there is more foreknowledge than necessary to produce completely on order, this extra foreknowledge can be used to smooth the arrival pattern to realize a further reduction of the inventory. For the case of complete foreknowledge, the system is equivalent to a queuing system with constant service rate. It is the reverse system. See figure 2.

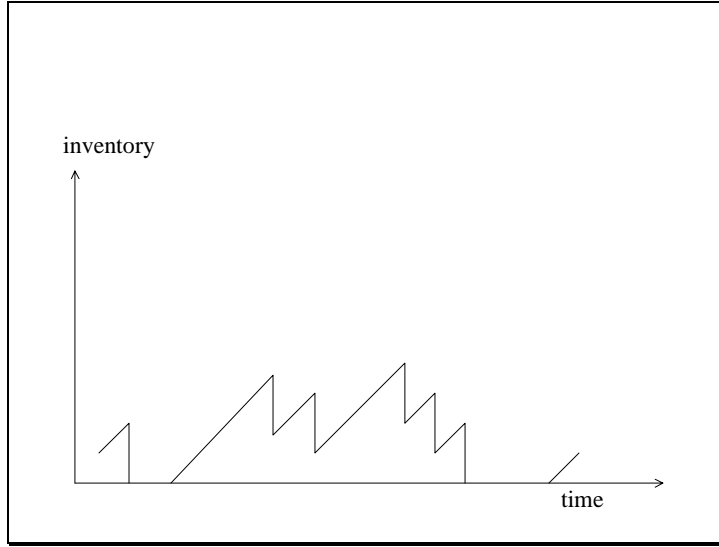


Figure 2: Reverse queue

In a queue with constant service rate the work in the system increases with jumps (due to arriving customers) and decreases gradually. Here the inventory increases gradually and decreases with jumps when the anticipated orders are delivered. The inventory pattern (as a function of $(+\infty, -\infty)$) is identical to the work content pattern in a queue with constant service rate.

Suppose the arrival pattern is Poisson with arrival rate 0.9. In case of complete foreknowledge it is possible to accept all orders. The resulting inventory is equal to

$$(1/2) \lambda / (1 - \lambda).$$

The analysis is well known in queueing theory, but for sake of completeness, the derivation is added in appendix 1. For $\lambda = 0.9$, the expression is equal to 4.5. For $\lambda = 1$, a fraction of 0.1 of the demand may be skipped. To reduce the inventory as much

as possible, the pattern of accepted orders should be made as flat as possible. It is optimal to smooth the pattern of accepted orders by introducing some critical level C such that new orders are refused if acceptance leads to more inventory than C units. Lemma 5 gives the result we need here.

Lemma 5

Consider a queue with Poisson arrival pattern ($\lambda < 1$) and constant service rate (1). All order sizes are equal to 1 and customers are serviced fifo. Arriving customers may be accepted or not. The optimal acceptance policy is of the critical level type. This means that there is a critical level C such that customers are accepted as long as the workload is smaller than C .

Proof

See appendix 2.

Table 8 gives some simulation results for $h = 60$, to show the effect of smoothing the arrival pattern.

Table 8 The influence of smoothing the arrival pattern (simulation runs of 10,000 time units)

Q = 1, B = 0					
$\lambda = 1, h = 60$					
C	#accepted	#skipped	#notaccept	Util.rate	Inventory
5	9021	792	90	0.903	1.901
6	9122	551	230	0.913	2.155
10	9302	106	495	0.930	2.836
∞	9329	0	574	0.933	2.967

To understand the effect of foreknowledge it is useful to make the following division:

- $h \leq B_p$:
Here the policy that is optimal for the case without foreknowledge can be applied. Acceptance and production may both be based on the (aggregate) variable $V(t)$. The bound B_p is defined so that applying the simple policy $(0, B_p)$ leads to utilization rate ρ .
- $B_p < h \leq h_p$:
Here it is not possible yet to produce completely on order. It is possible to use the simple aggregate policy as a starting point, but the production part of it has to be adapted to prevent stock-outs (slack checking). The forecast horizon h_p is defined so that pure make-to-order leads to utilization rate ρ .
- $h > h_p$:
Here it is possible to produce completely on order. Reduction of the average inventory can be realized by smoothing the arrival pattern.

A theoretically interesting question is whether combinations of smoothing and slack checking may be useful. The results presented below suggest that such combinations may be useful indeed.

Table 9 Combining smoothing with free inventory (simulation runs of 10,000 time units, three different random number generators)

Q = 1, $\lambda = 1, h = 33$		Random nmbr gen.					
C	B		#accepted	#skipped	#notacct	Util.rate	Inventory
∞	0	1	9100	0	803	0.910	2.055
∞	0	2	9193	0	670	0.920	2.150
∞	0	3	9189	0	753	0.919	2.114
∞	1	1	9114	0	789	0.912	2.089
∞	1	2	9212	0	651	0.922	2.200
∞	1	3	9193	0	749	0.920	2.123
7	1	1	9085	129	689	0.909	2.015
7	1	2	9190	107	566	0.919	2.140
7	1	3	9168	83	691	0.917	2.065
7.8	1	1	9106	55	742	0.911	2.054
8.8	1	3	9189	22	731	0.919	2.111

Adding smoothing to a policy means that certain orders are refused while there is sufficient capacity for acceptance. The results show that for the arrival pattern generated by generator 3, the combination of smoothing ($C = 8.8$) and free inventory ($B = 1$) leads to (a little) lower inventory than the pure make-to-order policy. That means that there is no easy sample path reasoning to show that pure make-to-order is better than the combination with free inventory. The conjecture is in fact that the optimal policy for $h > h_p$ may prescribe such combinations. This point is not investigated here further.

5 Conclusions and remarks

This paper investigates the effect of foreknowledge on the average inventory that is necessary to realize a certain pre-specified utilization rate in a capacity restricted production-inventory system where all accepted orders have to be delivered in time. Foreknowledge is modeled as a (constant) customer order leadtime.

The results show that if the foreknowledge does not exceed the buffer size (B_p) that is necessary to realize the utilization rate ρ , the foreknowledge affects the optimal policy in a trivial way. The optimal policy remains the same, but instead of the inventory itself, the virtual inventory has to be used. A customer order leadtime of h time units leads to a reduction of the inventory of $(1-\rho) \cdot h \cdot R$ (R the production speed). The proof is given for the case of order size 1 and Poisson arrival pattern, but the proof shows that for the more general case it is possible to derive similar results. The buffer size B_p that is required increases with ρ . In reality it is generally attractive to have a high utilization rate. In such cases it is reasonable to estimate the effect of foreknowledge as $(1-\rho) \cdot h \cdot R$.

Further research is necessary for the case of partial foreknowledge (customer order leadtimes that differ from each other), and for the case with some capacity flexibility (the possibility of using overtime for instance). Backlogging is treated in a subsequent paper, in combination with positive throughput time and more products (see Wijngaard[18]).

For cases with more foreknowledge the results confirm that no easy theoretical results may be expected. A further investigation of the possibility to combine a certain smoothing of the arrival pattern with using free inventory seems to be of interest and can help in providing general insight in the effect of foreknowledge.

REFERENCES

- [1] Baker, K.R. , “ An Experimental Study of the Effectiveness of Rolling Schedules in Production Planning”, *Decision Sciences*, Vol. 8, pp. 19-27, 1977.
- [2] Baker, K.R. , “An Analysis of Terminal Conditions in Rolling Schedules”, *European Journal of Operational Research*, Vol. 7, pp. 355 - 361, 1981.
- [3] Baker, K.R. ; D.W. Peterson, “An Analytic Framework for Evaluating Rolling Schedules”, *Management Science*, Vol. 25, pp. 341-351, 1979.
- [4] Blackburn, J.D.; R.A. Millen, “Heuristic Lot-Sizing Performance in a Rolling Schedule Environment”, *Decision Sciences*, Vol. 11, pp. 691-701, 1980.
- [5] Blackburn, J.D.; R.A. Millen, “The Impact of a Rolling Schedule in a Multi-Level MRP-System”, *Journal of Operations Management*, Vol. 2, pp. 125-135.
- [6] Blackburn, J.D.; D.H. Kropp; R.A. Millen, “A Comparison of Strategies to Dampen Nervousness in MRP Systems”, *Management Science*, Vol. 32, pp. 413 - 429, 1986.
- [7] Buzacott, J.A.; J.G. Shantikumar, “Stochastic Models of Manufacturing Systems”, Prentice Hall, 1993.
- [8] Buzacott, J.A.; J.G. Shantikumar, “Safety Stock vs Safety Time in MRP Controlled Production Systems”, *Management Science*, Vol. 40, pp. 1678 – 1689, 1994.
- [9] De Bodt, M.A.; L.N. Van Wassenhove, “Cost Increases Due to Demand Uncertainty in MRP Lot Sizing”, *Decision Sciences* Vol. 14, pp. 345-361, 1983.
- [10] Graves, S.C., “Safety Stocks in Manufacturing Systems”, *Journal of Manufacturing and Operations Management*, Vol. 1, pp. 67 – 101, 1988.

- [11] Hariharan, R.; Zipkin, P., “Customer-order Information, Leadtimes, and Inventories”, *Management Science*, Vol. 41, pp. 1599 – 1607, 1995.
- [12] Striekwold, P., “A Quantitative Comparison of Production and Inventory Control Concepts”, PhD-thesis University of Twente, 1988.
- [13] Van der Veen, B. “Safety Stocks: An Example of Theory and Practice in OR”, *European Journal of Operational Research*, Vol. 6, pp. 361 – 367, 1981.
- [14] Van der Veen, B. “Safety Stocks and the Paradox of the Empty Period”, *European Journal of Operational Research*, Vol. 16, pp. 19 - 33, 1984.
- [15] Vollmann, T.E.; W.L. Berry; D.C. Whybark, “Manufacturing Planning and Control Systems”, McGraw-Hill, 1997.
- [16] Whybark, D.C.; J.G. Williams, “Material Requirements under Uncertainty”, *Decision Sciences*, Vol. 7, pp. 595 – 606, 1976.
- [17] Wijngaard, J.; J.C. Wortmann, “MRP and Inventories”, *European Journal of Operational Research*, Vol. 20, pp. 281 - 293, 1985.
- [18] Wijngaard, J. “The Effect of Foreknowledge of Demand in Case of a Restricted Capacity: The Single-Stage Case with Backlogging. Research Report 02A16, University of Groningen (SOM), 2002.

APPENDIX 1

Consider a queue with Poisson arrival pattern ($\lambda < 1$) and constant service rate (1). All order sizes are equal to 1. The average work in the system can be determined as the quotient of the average surface of a busy period and the average length of the interval between two subsequent transitions to an empty system (recurrence to state 0).

Define $f(x)$ as the expected “cost” until the end of the busy period if the work load is now equal to x . Let $c(x)$ be the cost rate. By choosing $c(x) = 1$, $f(x)$ counts the length of the busy period. By choosing $c(x) = x$, $f(x)$ counts the surface of the busy period. The following equality holds for $f(x)$ and ε some small positive number:

$$f(x) = \varepsilon \cdot c(x) + (1 - \lambda \cdot \varepsilon) \cdot f(x - \varepsilon) + \lambda \cdot \varepsilon \cdot f(x - \varepsilon + 1),$$

neglecting the possibility of more than one arrival on the interval $(0, \varepsilon)$.

For $\varepsilon \rightarrow 0$, this leads to the following differential equation:

$$f'(x) = c(x) + \lambda \cdot (f(x+1) - f(x)). \quad (*)$$

For $c(x) = 1$, the solution we need of this equation is of the following form:

$$f(x) = \alpha \cdot x.$$

Substituting this in equation (*) leads to $\alpha = 1/(1-\lambda)$.

For $c(x) = x$, the solution we need is of the form:

$$f(x) = \beta \cdot x + \gamma \cdot x^2.$$

Substituting this in equation (*) leads to $\beta = (1/2) \cdot \lambda / (1 - \lambda^2)$ and $\gamma = (1/2) / (1 - \lambda)$.

The expected surface of a busy period is equal to:

$$\beta \cdot 1 + \gamma \cdot 1 = (1/2) / (1 - \lambda)^2.$$

The expected length of the interval between two subsequent transitions to an empty system (recurrences to state 0) is equal to:

$$\alpha \cdot 1 + 1/\lambda = 1/(\lambda \cdot (1 - \lambda)).$$

This implies that the average workload is equal to:

$$(1/2) \cdot \lambda / (1 - \lambda).$$

APPENDIX 2

This appendix considers the same system as appendix 1. But here the order acceptance decision is investigated. The cost rate of having workload x is assumed to be equal to $c(x)$, with $c(x)$ increasing in x . The reward for each accepted order is equal to r . It is proved that the optimal acceptance policy is of the type: accept if the workload is smaller than some critical level C . This implies also that a pre-specified ρ can be realized with minimal average workload by such a critical level policy. That is because for each pre-specified ρ the value of r can be chosen such that the optimal policy leads to a utilization rate ρ (compare lemma 3).

To prove that the optimal policy for the reward problem is of the critical level type, policy iteration is used. Let g be the minimal average cost. Define the value function $v(\cdot)$ as the minimal expected value of the “cost” until the first recurrence to 0, with the “cost” consisting of two elements:

- the cost rate (workload x gives cost rate $x - g$)
- the negative cost for accepting orders (each acceptance gives a negative cost r)

The term $-g$ is added to realize that the expected cost until recurrence is equal to 0. Using the same approach as in appendix 1 leads to the following expression for $v(\cdot)$:

$$v'(x) = x - g + \lambda \cdot \min\{v(x+1) - r, v(x)\} - \lambda \cdot v(x).$$

Since $v(x+1) - v(x)$ is equal to the expected cost until the first visit to x , starting in $x + 1$, and because of the fact that the cost is increasing in x , it follows that $v(x+1) - v(x)$ is increasing in x . That means that $v(x+1) - r < v(x)$ implies $v(y+1) - r < v(y)$ for all $y < x$. This proves that the optimal policy is of the critical level type.